



VBRC

Viral Bioinformatics Resource Facility Annotation, Curation, and Collaboration

VBRC

- Annotation
 - Computationally-derived information describing the genes and genome of a particular viral sequence
 - VBRC Database
- Curation
 - Human-derived information describing the genes and genome of a particular viral sequence
 - Knowledge Database (KDB)
- Collaboration
 - Extension of the KDB into a multi-user collaborative application
 - Collaborative Virtual Environment (CVE)

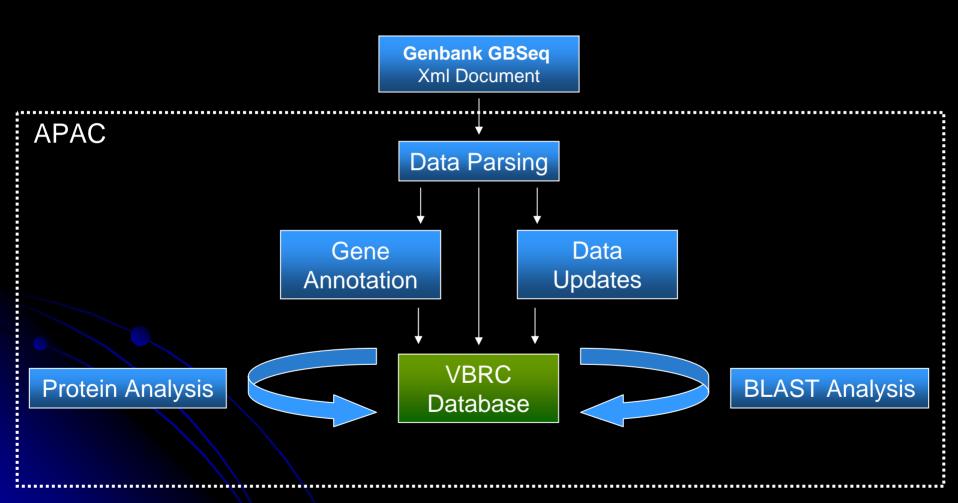
Annotation Pipeline

APAC

APAC: Automated Parse Annotate Curate

- A java application
- Series of modules that automate:
 - Data parsing
 - Annotation of incomplete genome records
 - Genome record updates
 - Gene analysis
 - Computational pipeline

Overview of pipeline for data acquisition from Genbank to VBRC database



High Performance Computing Tools for Annotation

- Computationally-intensive Bioinformatics analyses
 - Linux Clusters
 - Grid Computing
- Similarity searching
 - SS-Wrapper
 - NCBI BLAST, WU BLAST, FASTA, PC_SCAN, HMMPFAM...
- Refinement of genome-scale multiple sequence alignments
 - GenAlignRefine
- Poxvirus gene prediction
 - Sequence Signals (Promoter prediction, Glimmer)
 - Similarity (BLAST and HMMPFAM)
 - Comparative analyses (Orthologs and Gene synteny)

Curation

KDB

VBRC Knowledge Database Curation

- Mini review of available structure-function information
 - Human-curated database based on the literature
- Bibliographic information
- Available scientific resources
 - clones, mutants, and antibodies
- Empirically-derived properties
 - MW, pl . . .
 - Post-translational modifications
 - Expression
- Functional Assignments
 - Gene Ontology controlled vocabulary
 - Molecular function
 - Biological Process
 - Cellular component
 - Virulence Ontology

Three-Tiered Curation Process

- 1) UAB/UVic full-time curators
 - B.S., M.S., Ph.D.-level personnel
- 2) Editorial Board
 - Domain experts
 - Review records generated from tier 1
- 3) Scientific Community
 - Feedback

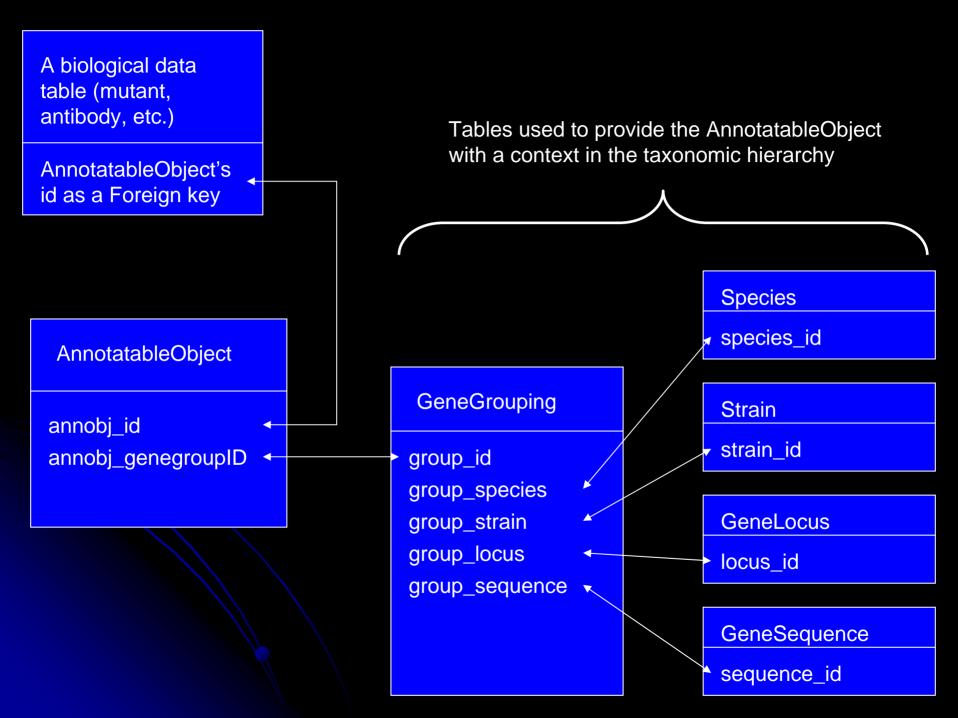
KDB Curation Application

- Provides a systematic, controlled interface for entering curation data
 - Jason Pokorny, MCW
 - Initial and ongoing development
 - Don Dempsey, UAB
- VB.Net Application
 - Web interface



KDB Data structure

- Biological Objects (BO)
 - Antibodies, Gene Homologs, Gene Interactions, Functions
 - Mapped to references and other information
- Annotatable Objects (AO)
 - Maps BO to sequence
 - Maps to a particular level in the taxonomic hierarchy
 - species, strain, gene, gene sequence



AnnotatableObject

annobj_id

An example of how a biological data table (Antibody, in this example) is associated to the AnnotatableObject table and the References table.

Antibody

antibody_id antibody_annobjID antibody_refID antibody_name antibody_type

References

reference id

AntibodyAppSel

antibodyappsel_id antibodyappsel_antilD antibodyappsel_appID

AntibodyAppSel associates an antibody and an application.

AntibodyApp

antibodyapp_id antibodyapp_refID antibodyapp_name

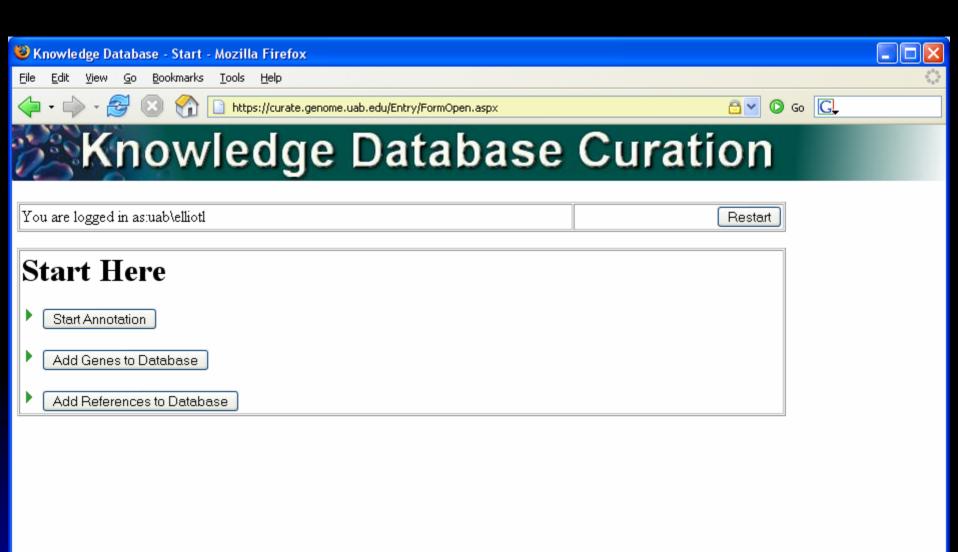
Affinity purification
Immunofluorescence
Neutralizing

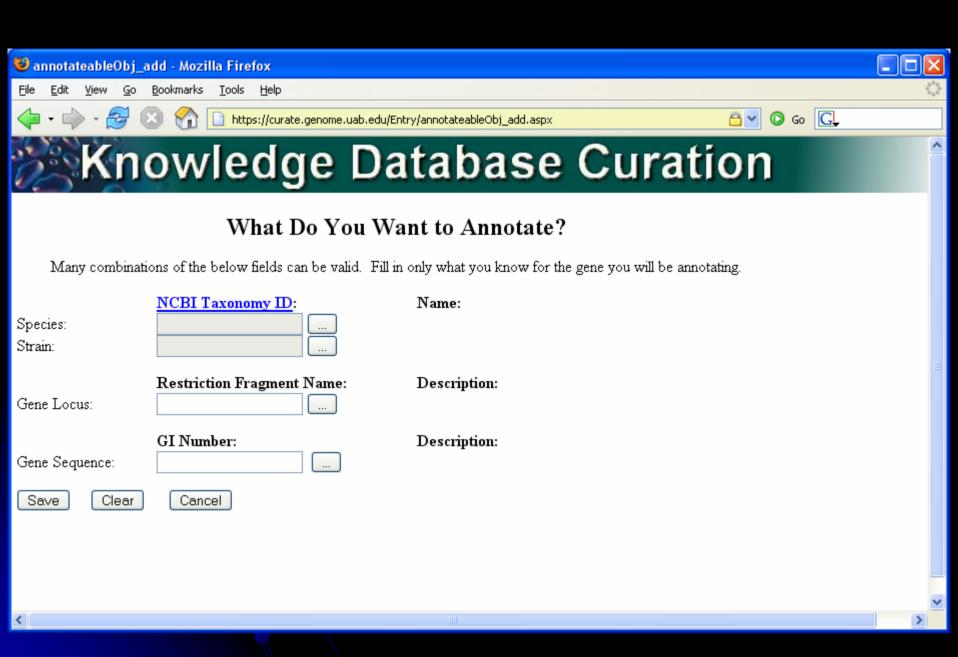
Curation Starting Point: The Reference Genome

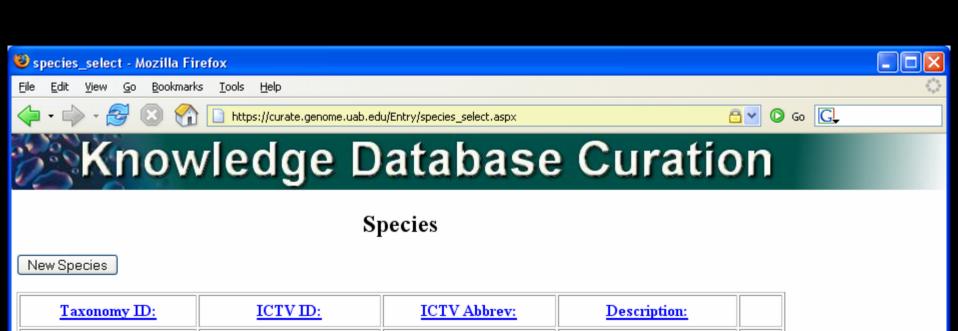
- Viral strain representing a "prototype" for the species
 - Complete genomic sequence
 - Extensive literature references
 - "Laboratory strain" used for wet-bench work
- Curation of remaining strains
 - Link prototype to all remaining strains within a species
 - Link other strains to prototype
 - Curate differences between strains
 - Extrapolate information and evidence from one strain to another

Source of Curation Data

- Fields Virology (4th edition, 2001)
- Literature searches
- Literature mining
- Domain experts







Vaccinia Virus

S

IVACV

Go



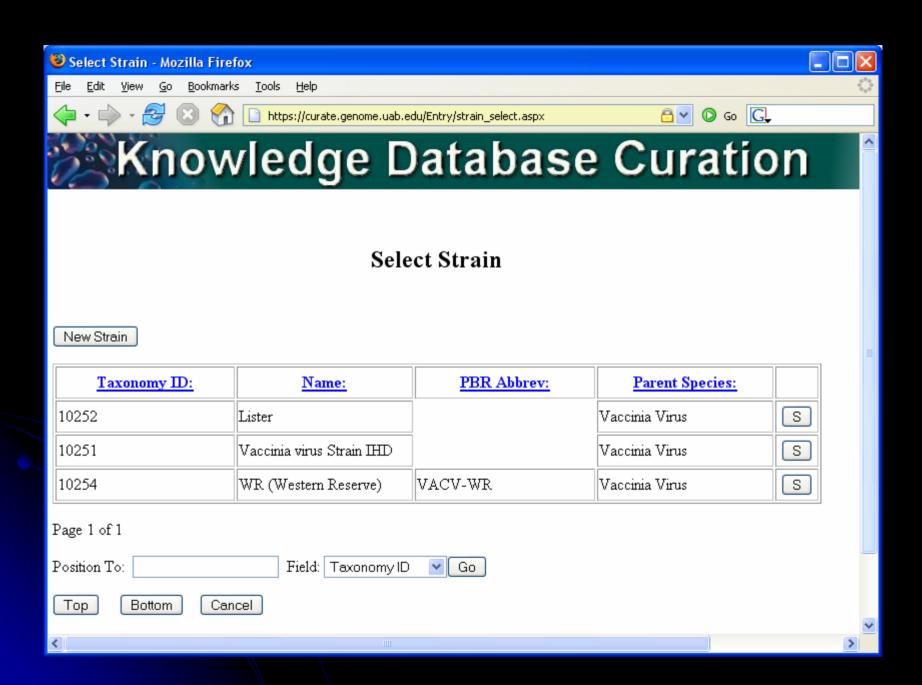
Cancel

00.058.1.01.001

10245

Тор

Bottom



annotateableObj_select - Mozilla Firefox



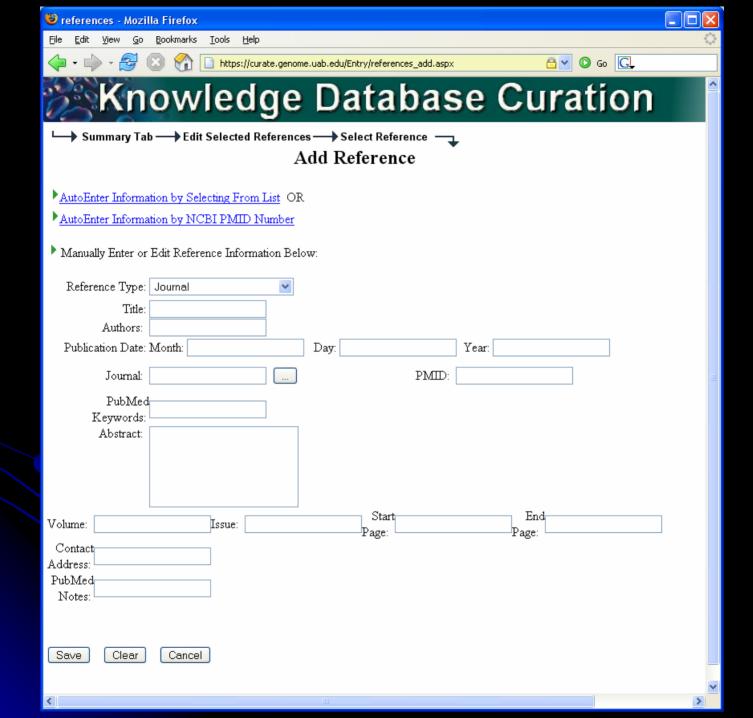
<u>File Edit View Go Bookmarks Tools Help</u>

Knowledge Database Curation

	midage Batabade Garation			
Actions: Create	lew Annotatable Object Quit Annotating			
Position to 'Locus	- name': Go Top Bottom			
Subset: Species - I	ame v = v Subset			
	Page 1 of 33 Next >>			
	Annotatable Objects: des-Name Species-taxID Species-PBR Strain-Name Strain-taxID Locus-name cous-description Sequence-GI Sequence-gene Sequence-product Sequence-definition			
Nucleoprotein	SPECIES: Name: Lake Victoria marburgvirus taxID: 11269 ICTV: MBG PBR: MBG STRAIN: Name: Musoke taxID: 33727 PBR: MUS LOCUS: Name: MARV NP Description: Marburg nucleoprotein SEQUENCE: GI:60624 Gene: 18 Product: nucleoprotein Definition: nucleoprotein	s		
VP35 Polymerase Complex Protein				
Delta peptide cleavag product of sGP	SPECIES: Name: Zaire ebolavirus taxID: 186538 ICTV: ZEBOV PBR: ZEBOV STRAIN: Name: Mayinga taxID: 128952 PBR: MAYI LOCUS: Name:delta peptide Description: cleavage product of sGP	S		
second small non-structural secreted glycoprotein	SPECIES: Name: Zaire ebolavirus taxID: 186538 ICTV: ZEBOV PBR: ZEBOV STRAIN: Name: Mayinga taxID: 128952 PBR: MAYI LOCUS: Name:ssGP Description: Second small non-structural secreted glycoprotein ssGP SEQUENCE: GI:10313996 Gene: 11 Product: second small non-structural secreted glycoprotein Definition: second small structural secreted glycoprotein	S		
/P35 polymerase omplex protein	SPECIES: Name: Zaire ebolavirus taxID: 186538 ICTV: ZEBOV PBR: ZEBOV STRAIN: Name: Mayinga taxID: 128952 PBR: MAYI LOCUS: Name:VP35 Description: Polymerase Complex Protein SEQUENCE: GI:10313992 Gene: 2 Product: <none> Definition: VP35 Polymerase Complex protein</none>	S		
age 1 of 3 Next >	>			
ort By: Species - na	me Go Field: Species - name Go			
Top Bottom				

Knowledge Database Curation SPECIES: Name: Zaire ebolavirus taxID: 186538 ICTV: ZEBOV PBR: ZEBOV Currently STRAIN: Name: Mayinga taxID: 128952 PBR: MAYI Annotating: LOCUS: Name: ZEBOVgp1 Description: np (nucleoprotein Change SEQUENCE: GI:10314000 Gene: 3 Product: <NONE> Definition: nucleoprotein Identification Homology Qual. Properties Quant. Properties Summary Function Reagents Interactions Manual Save The nucleoprotein (NP) of Zaire Ebola is involved in encapsidation of viral RNA. It is Mini-Review: |comprised of a hydrophobic N-terminal domain involved in RNA binding and an acidic hydrophilic encapsidation of viral RI Short Description (max 150): Long Functional Name (max 80): nucleoprotein Short Functional Name (max 30): PBR (poxvirus.org) ID for ZEBOV-MAYI 001 gene/virus: Miscellaneous notes: Selected References: Add Reference Delete Selected References PMID: Date: Authors: Title: Journal: Edit Del Niikura M. Ikegami T. Detection of Ebola viral antigen by

11526161	Sep 2001	Saijo M, Kur	enzyme-linked immunosorbent assa	Microbiol	E	
Notes: Mono	oclonal Antibod	ies, 3-D and 2-11G recog	nize conformational epitopes at C-term	inus.		
12191476	Aug 2002	Huang Y, Xu L, Sun Y, Nabel GJ	The assembly of Ebola virus nucleocapsid requires virion-associated	Mol Cell	E	
Notes: Nucle	eoprotein is O-I	inked glycosylated.				
12522044	Jan 2003	100 CC 200 CC 20	Analysis of linear B-cell epitopes of the nucleoprotein of ebola vir	Clin Diagn Lab Immunol	E	
Notes: MAb	s to the C-term	inal region of ebolavirus N	P have different subtype specificities.			



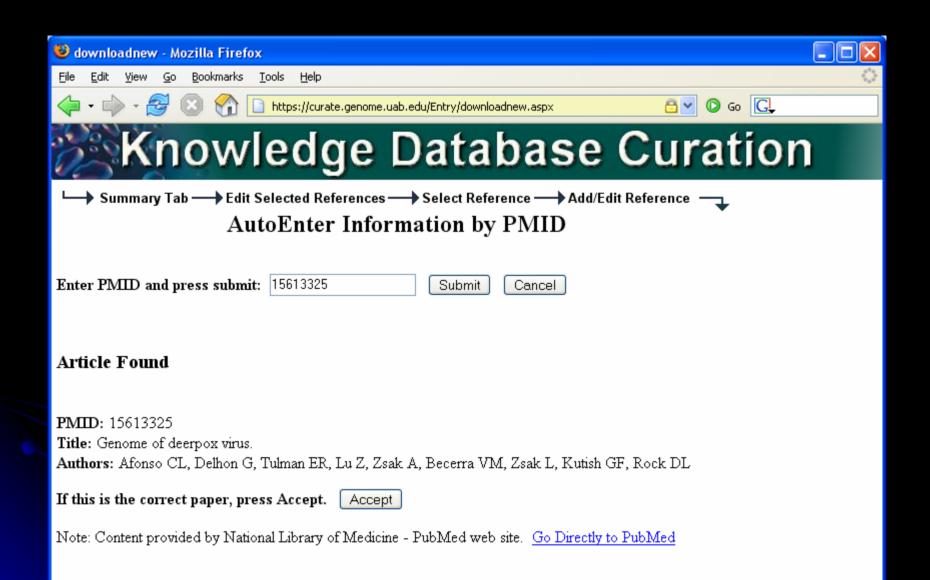
Knowledge Database Curation

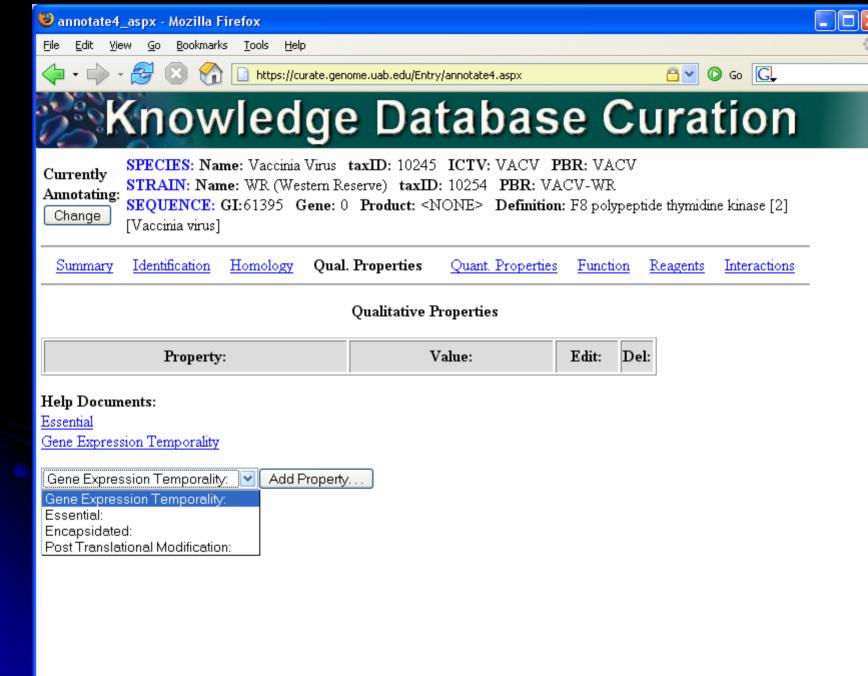
Summary Tab —	→ Edit Selected References —
	References

Actions: Create Reference	Cancel	
Position to 'PubDate':	Go	Top Bottom
Subset: PMID =	Subset	

Page 1 of 3 Next >>

PMID:	<u>Title:</u>	Authors:	PubDate:	Journal:	
4712965	Phosphoprotein component of vaccinia virions.	Rosemond H, Moss B	Jun 1973	J Virol	S
1169849	Biogenesis of poxviruses: synthesis and phosphorylation of a	Pogo BG, Katz JR, Dales S	Apr 1975	Virology	S
467437	Phosphorylation in vivo of a vaccinia-virus structural prote	Sagot J, Beaud G	Jul 1979	Eur J Bioch	S
6894655	Purification and characterization of a superhelix binding pr	Kao SY, Ressner E, Kates J, Bauer WR	Jun 1981	Virology	S
7269243	Unit Complex of vaccinia polypeptides linked by disulfide br	Ichihashi Y	Aug 1981	Virology	S
6890583	A phosphorylated basic vaccinia virion polypeptide of molecu	Hiller G, Weber K	Nov 1982	J Virol	S
6319738	Mapping of a gene coding for a major late structural polypep	Wittek R;Hanggi M;Hiller G;	Feb 1984	J Virol	S
3856886	One hundred base pairs of 5' flanking sequence of a vaccinia	Bertholet C, Drillien R, Wittek R	Apr 1985	Proc Natl A	S
3617504	Biosynthesis and phosphorylation of vaccinia virus structura	Kao SY, Bauer WR	Aug 1987	Virology	S
2835495	Sequence and transcriptional analysis of the vaccinia virus	Schmitt JF, Stunnenberg HG	Jun 1988	J Virol	S





Knowledge Database Curation

SPECIES: Name: Vaccinia Virus taxID: 10245 ICTV: VACV PBR: VACV Currently STRAIN: Name: WR (Western Reserve) taxID: 10254 PBR: VACV-WR Annotating: SEOUENCE: GI:61395 Gene: 0 Product: <NONE> Definition: F8 polypeptide thymidine kinase [2] Change [Vaccinia virus] Identification Qual. Properties Quant. Properties Function Summary Homology Interactions Reagents Localization ▶ Biological Function Molecular Function Cellular Localization Add Localization Delete Selected Localizations Number of GO Number of Localization: Edit: Notes: References: Terms: **Biological Function** Delete Selected Functions Add Biological Function Number of Number of GO **Biological Function:** Edit: Notes: References: Terms:

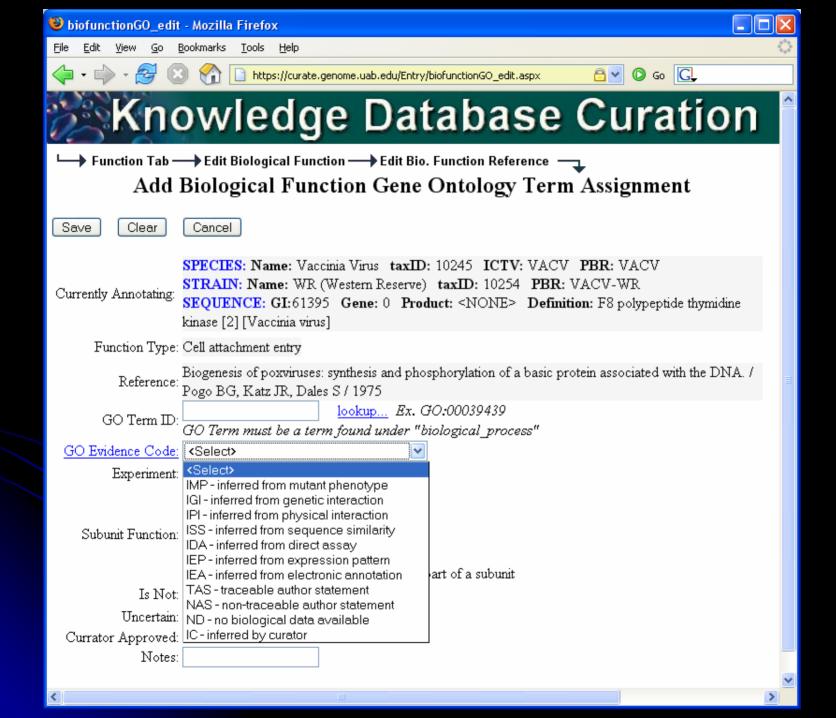
Molecular Function

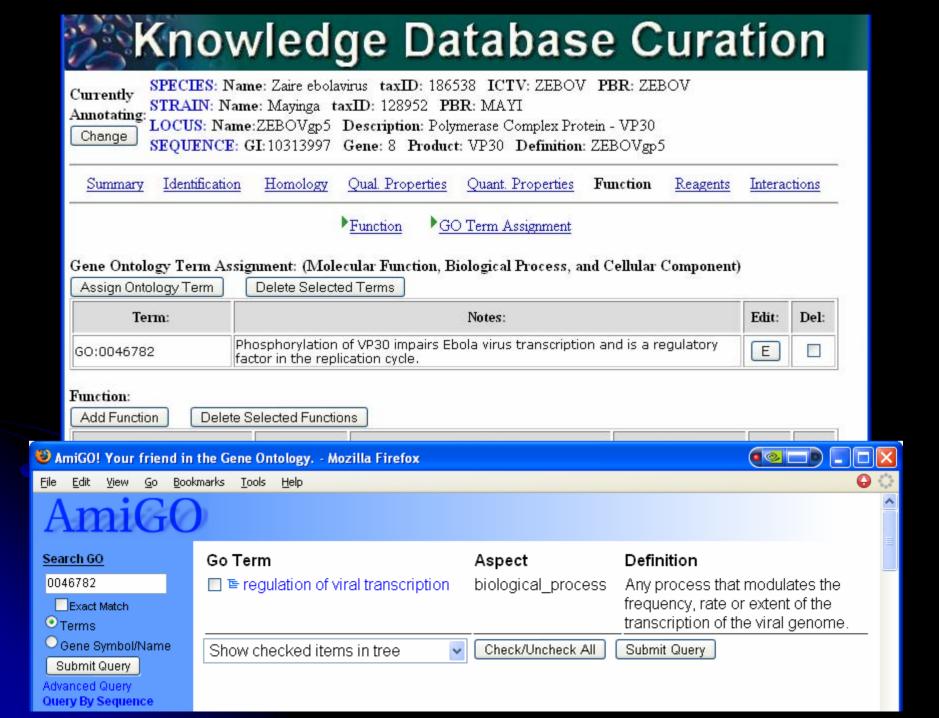
Add Molecular Function Delete Selected Functions

Molecular Function:	Notes:	Number of References:	Number of GO Terms:	Edit:	Del:	
---------------------	--------	-----------------------	------------------------	-------	------	--

Del:

Del:





Knowledge Database Curation

Currently Annotating: Change SPECIES: Name: Vaccinitation STRAIN: Name: WR (SEQUENCE: GI:61395 [Vaccinia virus]	Western Reserve)	taxID: 1025	54 PBR : VA	CV-WR		e kinase	[2]
Summary Identification Homol	ogy Qual Prop	erties Quant	. Properties	Function	Reagents	Interac	tions
Ar	tibodies Mo	dified Virus	Expression	Vectors			
Primary Antibodies to ZEBOVgp1F	alse:						
Add Antibody Delete Selected	Antibodies						
Antibody:	Type:		Create Method:		Edit:	Del:	
3-1E	monoclonal		mouse			E	
Ref: PMID: 12522044 Date: Jan 2003	, Saijo M, Kurata T, Kurane I, Morika						
Contact:			Notes:				
3-3D and 2-11G	monoclonal		mouse			E	
Ref: PMID: 11526161 Date: Sep 200	1 Authors: Niikur	a M, Ikegami T	, Saijo M, Kur	ane I, Mirar	nda ME, Mori		
Contact: Shigeru Morikawa	Notes:						
Modified Virus: Add Modified Virus Delete Sele	cted Virus						
Modified Virus Name	Туре:		Notes:		Edit:	Del:	
Recombinant - TK gene with 11kDa k					E		
Expression Vectors							
Add Vector Delete Selected Ve	ctor						
Expression Vector Type:				Notes:		Edit:	Del:

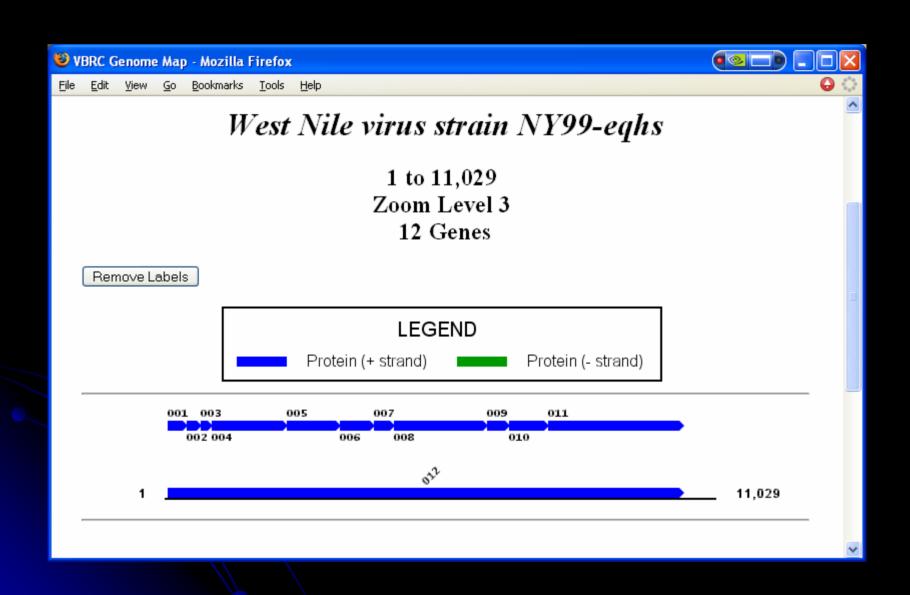
Identification and Correction of Errors

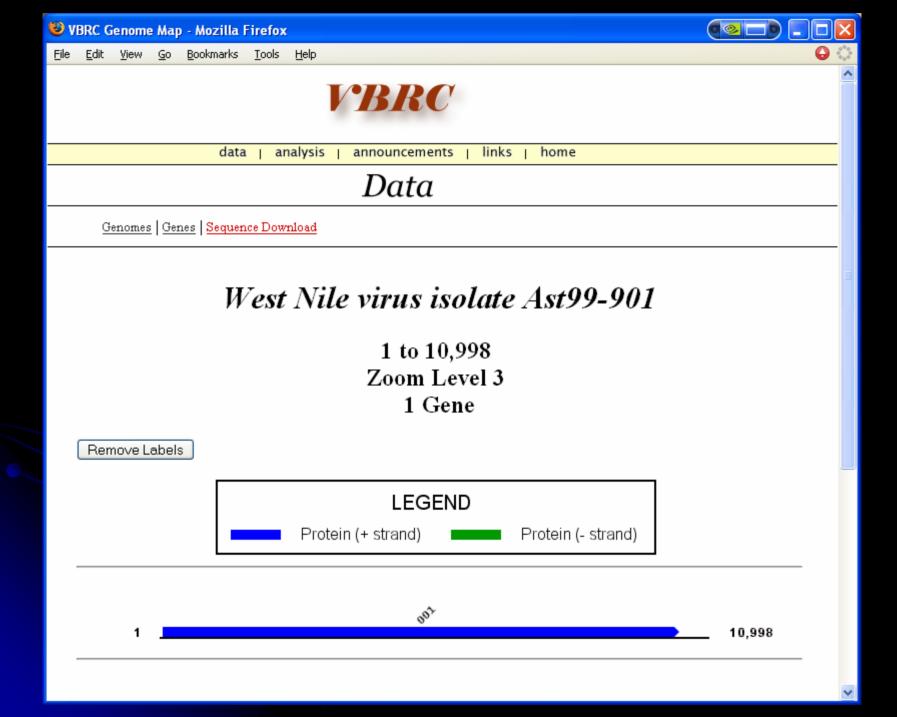
Sequence Errors

- Vaccinia Virus strain Copenhagen
 - Reported in the literature
 - Senkevich et al Virology V 300 p 296
 - Missing "c" at position 114378
 - NOTE: Error still present in the RefSeq record
- Vaccinia Virus strain Tian Tan
 - Identified during PBR annotation
 - Many sequence errors resulting in fragmented ORFs
 - UVic resequenced
- Ebola virus strain Zaire 1995
 - Identified during VBRC annotation
 - Extra "A" in the RNA editing region of the genome
 - Confirmed by S. Ibrahim, submitter

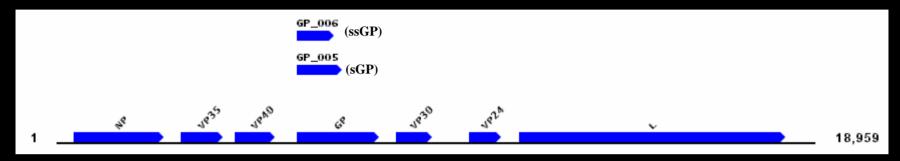
Annotation Errors

- Incomplete annotation
 - No gene prediction
 - Incomplete prediction

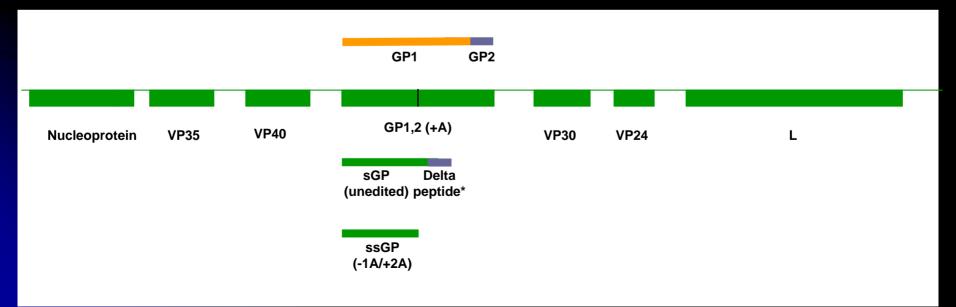




Zaire ebolavirus strain Mayinga CDS annotation



Zaire ebolavirus strain Mayinga VBRC annotation



Collaborative Virtual Environment (CVE)

Purushotham Bangalore Barrett Bryant Ziyang Xiong Ritu Arora

Department of Computer and Information Sciences University of Alabama at Birmingham (UAB)

Collaborative Knowledge Database

- CKDB
- Provides for collaborative interactions among KDB users within a single user interface

Requirements

- Support distributed and collaborative
 - Authoring/Editing
 - Curation
 - Information sharing and exchange
- Provide different interfaces/tools for different users through a web interface
- Promote interoperability and web-based linkage tools for data sharing among other BRCs

Different User Roles

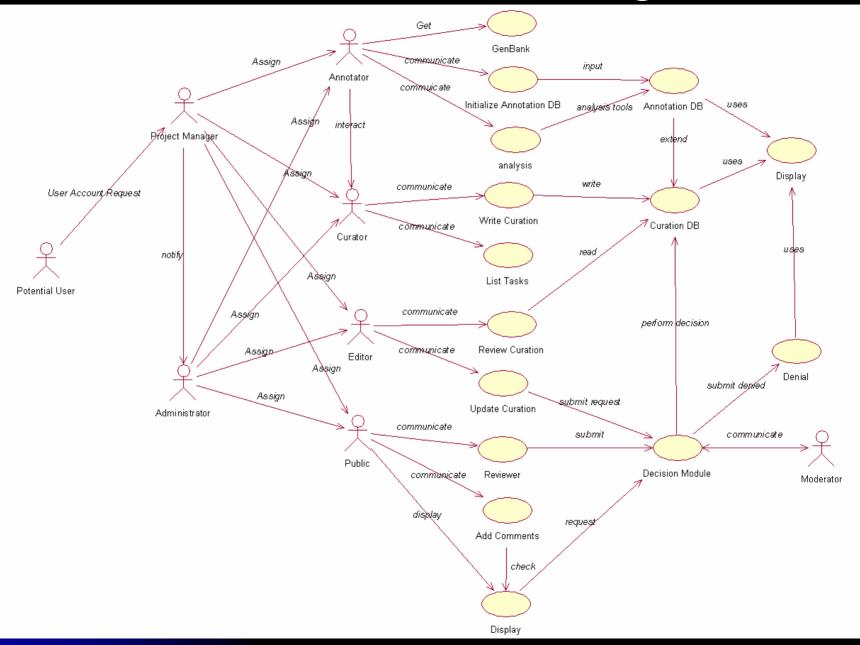
- Annotator
- Curator
- Moderator
- Editor
- Public
- System Administrator
- Project Manager
- Other BRC application/program
- Demo Role

Each role provides a different set of functionality and/or user interface

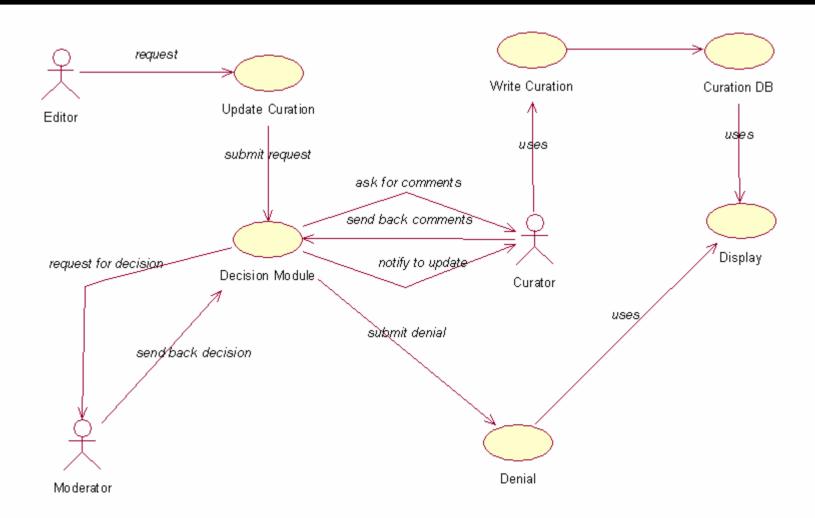
Mapping Roles With Services

- A user can have multiple roles
- Each user has to go through a registration process to request access to the portal
- At registration user can specify/request appropriate roles
- Portal administrator will approve user request and assign appropriate roles
- Portal uses user profiles to provide appropriate interfaces and access to required services
- Access control can be coarse grained (service level) or fine grained (data record level)

Overall Use Case Diagram



Use Case: Editor Change Request



Current Status

- Identification of different roles
- Identification of different services
- Use case analysis for each role and service
- Prototype portal with skeletal support for multiple roles leveraging KDB
- Implement collaborative services not currently provided in KDB